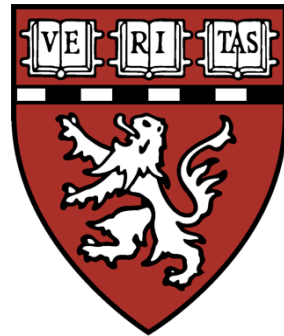


Spaced Repetition for Training Artificial Neural Networks

Area: 2. Neural Networks and Deep Learning (DL)

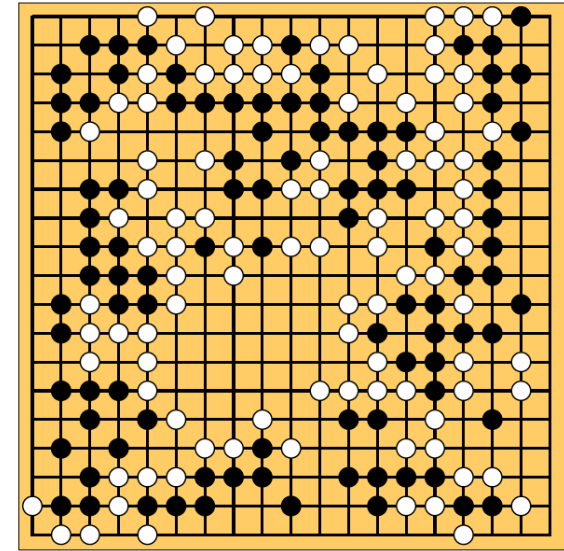
Hadi Amiri, PhD

<http://scholar.harvard.edu/hadi/>



Neural Networks

- Detect objects in images
- Play strategy games
- Predict medical events
- Translate text
- Etc.



GO Silver et al. (2016)

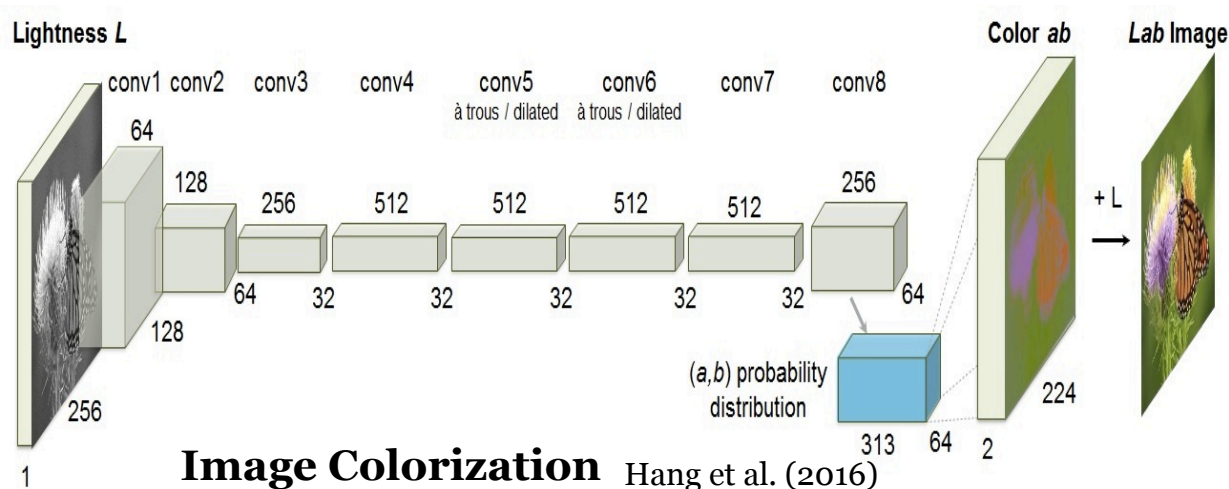
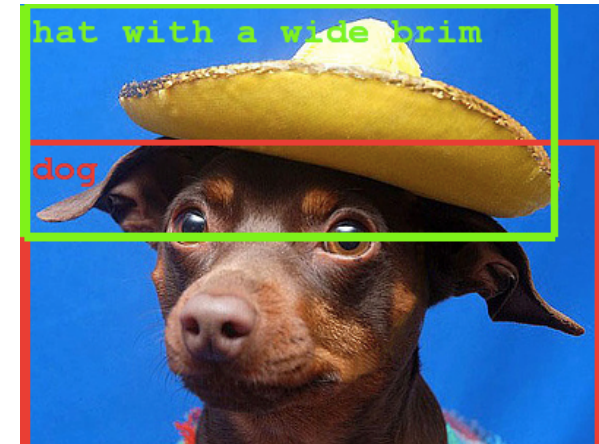


Image Colorization Hang et al. (2016)



Object Detection LeCun (1995)

Challenges

- Data hungry
- Hard-to-interpret
- Easily fooled
- Finicky to optimize
 - complex architectures
 - many hyperparameters
 - learning procedure
 - overfitting/underfitting
 - initialization, etc.
- Compute-intensive
 - long training time

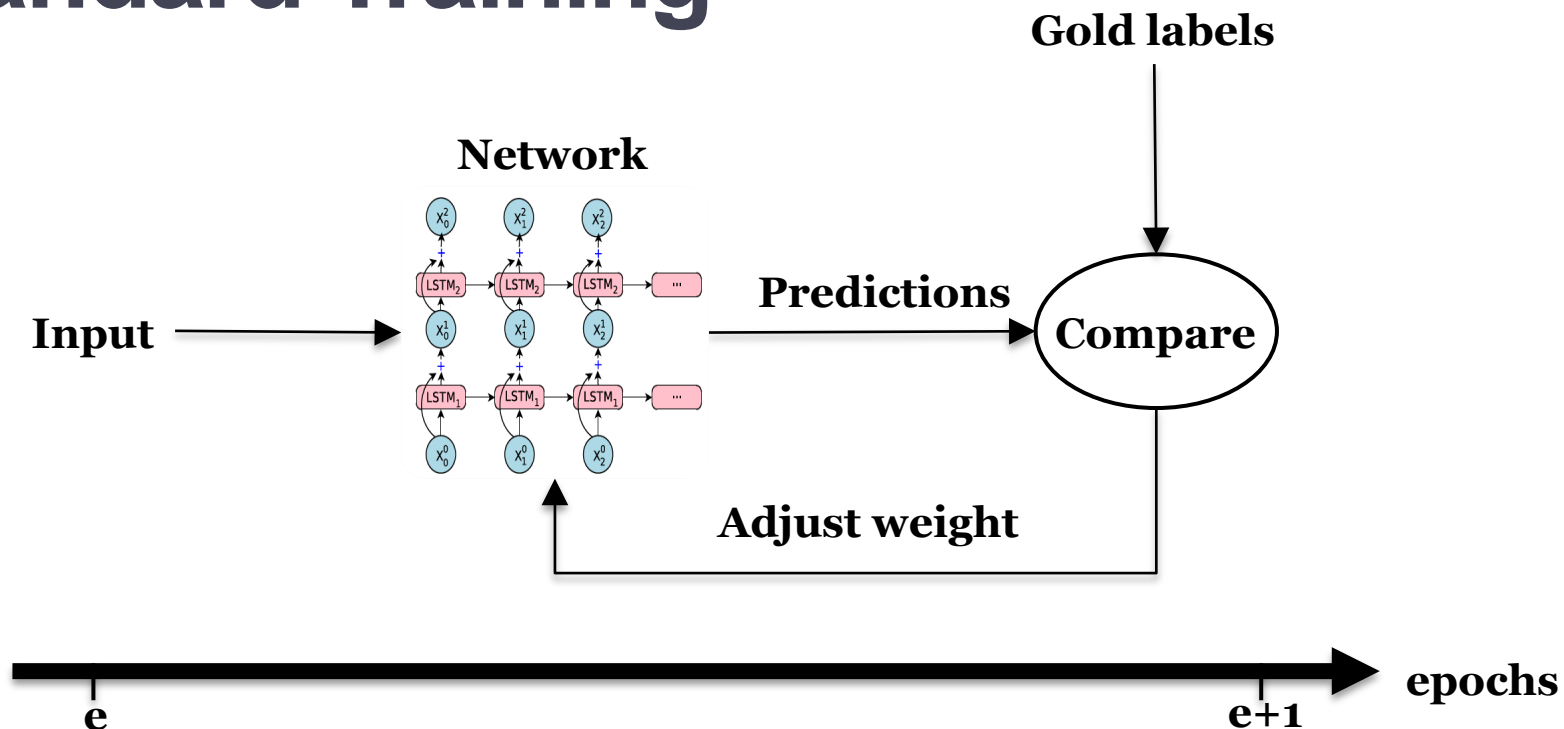
Challenges

- Data hungry
- Hard-to-interpret
- Easily fooled
- Finicky to optimize
 - complex architectures
 - many hyperparameters
 - learning procedure
 - overfitting/underfitting
 - initialization, etc.
- **Compute-intensive**
 - long training time
- **Cognitive Psychology**
 - Attention
 - Language use
 - **Learning & memory**
 - Perception
 - Problem solving
 - Creativity
 - Thinking

Amiri, et al. (NAACL 2018)

Amiri, et al. (EMNLP 2017)

Standard Training



Could result in long turnaround time depending on data size & network complexity



Neural MT Training time: 3 weeks on 100 GPUs

Goal and Solution

Develop **efficient** and **effective training paradigms** for neural models.

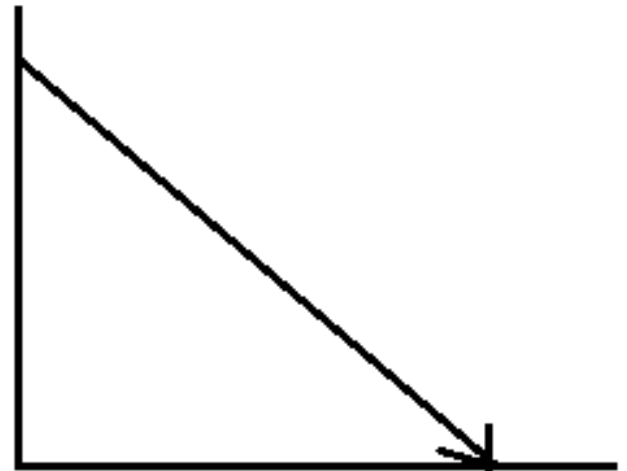
Efficient: Shorter training time

Effective: Comparable/better performance on test data

- Spaced Repetition
 - Spread reviews over time and increase time intervals between subsequent reviews of previously learned materials.

learning materials

Start BIG



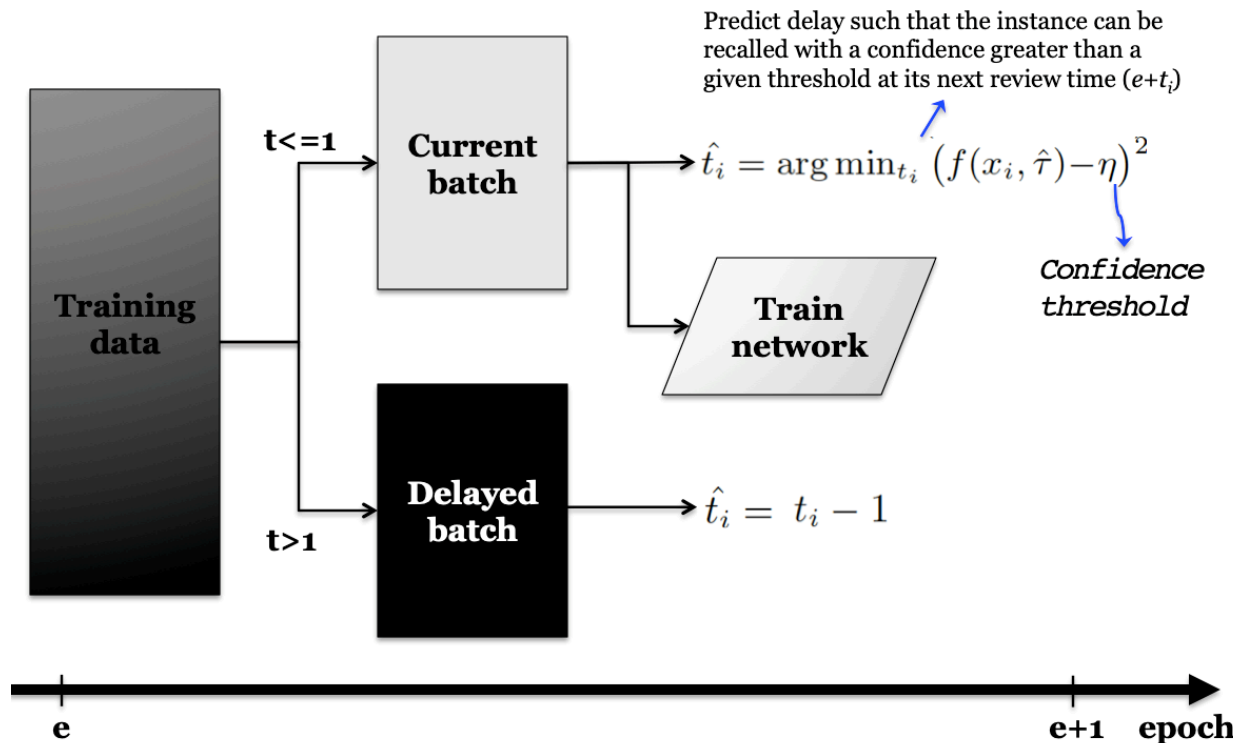
Spaced Repetition

- Recall Indicators

- Delay since last review of the item
- Difficulty of the item
- Strength of the human memory

Attributes of **item**

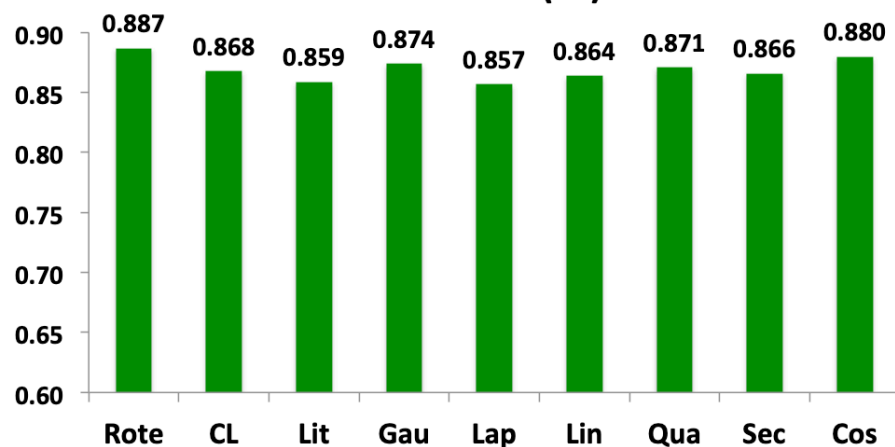
Attribute of **learner**



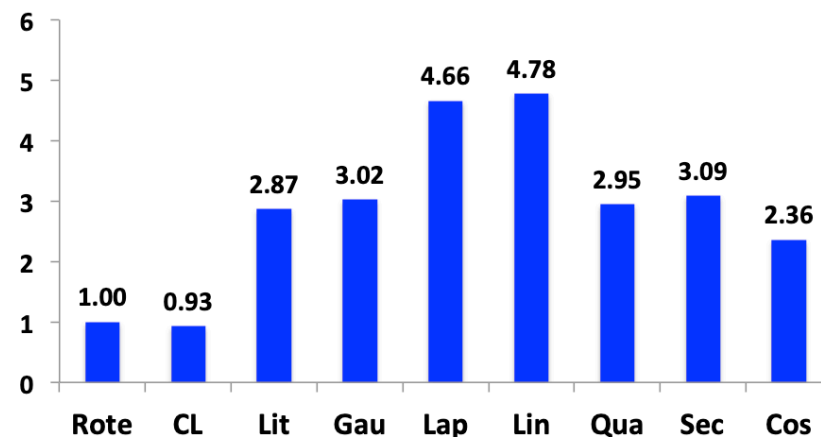
Effectiveness & Efficiency

Batch size = 1

Performance (F1)



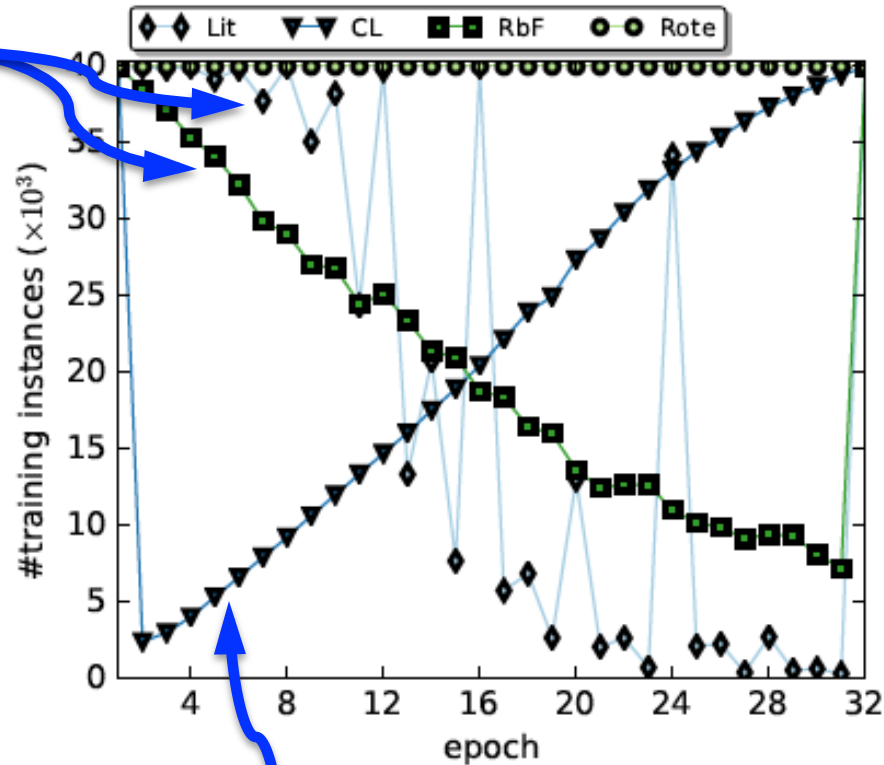
X-time Faster



- Spaced repetition kernels are **considerably faster** than standard training.
- The kernels show **comparable accuracy** to standard training.

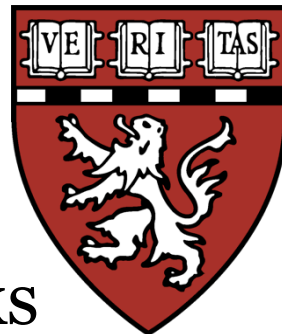
Efficiency

Spaced repetition
starts **BIG**



Curriculum learning
starts **SMALL**

Contributions



- Novel training paradigms for neural networks based on spaced repetition
- Model can be applied without modification to any neural network

Code & Data

- <https://scholar.harvard.edu/hadi/pubs>
- hadi.amiri@childrens.harvard.edu